

REMARKS

Applicants note that the claim amendments submitted with the September 30, 2009 Response and Amendment under 37 CFR §1.111, have been entered and forwarded to the Examiner for review, these have been considered by the Examiner, as discussed during the Interview with the Examiner on November 10, 2009 and referenced in the Examiner Interview Summary mailed November 17, 2009. Accordingly, Applicants' claim amendments in this Supplemental Response and Amendment under 37 CFR §1.111 reference the entered claim listing provided in the September 30, 2009 filing.

Reconsideration and withdrawal of the rejections set forth in the Office Action mailed April 30, 2009 are respectfully requested in view of this response and amendment. Claims 1-12, 14-17 and 19-21 are pending in this application. Claims 1-12 and 14-18 have been rejected.

Claims 1, 14, 15 and 19 have been amended for the sole reason of advancing prosecution. The claims have been amended to clarify the claimed subject matter, as requested by the Examiner and otherwise conform the claims to U.S. patent practice. Applicants, by amending any claims, make no admission as to the validity of any rejection made by the Examiner against any of these claims. Applicants reserve the right to reassert the original claim scope of any claim, in a continuing application. Support for the claims as amended appears throughout the specification, claims and figures as originally filed.

Claims 20 and 21 are newly presented. The subject matter of newly presented independent claim 20 is supported throughout the specification, claims and figures as originally filed, at least by original and amended claim 1 and page 14, line 1, et seq. of the original specification. The subject matter of newly presented dependent claim 21 is similarly supported, and at least by original and

amended claim 1 and page 15, line 11, et seq. of the original specification. It is respectfully submitted that the amendments do not introduce any new matter within the meaning of 35 U.S.C. §132.

In view of the following, further and favorable consideration is respectfully requested.

Applicants hereby fully reference and incorporate any applicable remarks and arguments made in the Response and Amendment filed September 30, 2009 in response to the Office Action mailed April 30, 2009. During the Interview conducted November 10, 2009, the Examiner indicated that the amendments of the Response filed September 30, 2009 overcame the objections to claims 1 and 18, and overcame the rejections under 35 U.S.C. §101; accordingly, only remarks regarding the prior art rejections are made below.

Interview Summary

Applicants gratefully appreciate the courtesies extended to Applicants' Representatives, during the Interview conducted on November 10, 2009. Discussion focused primarily on the independent claim. Applicants are grateful that the Examiner (A. RUIZ) and the Examiner's Primary Examiner (S. PANNALA) clarified their understanding and broadest interpretation of several terms of Applicants' claimed subject matter. From Applicants' Representatives' explanation of the Applicants' subject matter, the Examiners acknowledged that "the [the described subject matter] is a little different than some pioneers [in this art]" and that the disclosed subject matter is quite possibly patentable if the claims are amended somewhat. Applicants note that during the Interview, the Examiner indicated that she had reviewed Applicants' filed Response but not yet formally acted on

it, thus Applicants were encouraged to make further amendments to advance prosecution.

Applicants' Representatives gave an overview of the claimed subject matter respective to the real world application, discussing that the subject matter supports both "corpus analysis" and "retrieval" (i.e. keyword search, cluster browsing means). Within "corpus analysis," a first step (referred to as "narrow context discovery" in the specification, and performed *prior* to "document clustering" and "internal cluster structuring") is particularly novel in calculating a probability distribution indication of the frequency of occurrence of each term of a document and other terms of the document, calculating the entropy of each probability distribution, and selecting at least one of the probability distributions as a cluster attractor depending on the entropy value. Applicants' Representatives clarified these features with respect to the independent claim 1 ("method of determining cluster attractors *for a plurality of documents ... calculating, in respect of each term, a probability distribution [that is] indicative of the frequency of occurrence of one other term ... the respective frequency of occurrence of each other term, that co-occurs with said term in at least one of said documents...*"') and its distinguishing features over the prior art with the arguments discussed in the Response filed September 30, 2009. The relationship between independent method claims 1 and 15 and independent apparatus claims 14 and 19 was also discussed.

The Examiners requested clarification with regard to whether Applicants disclose specific formulas for the probability distribution or entropy calculation (the Examiners agreed such disclosure was provided), how many 'terms' may be in a 'document,' and for which terms the respective frequency of occurrence is calculated. Applicants' Representatives asked the Examiner for clarification responsive to Applicants' submission that Choi "order-rank[s] document clusters" and

therefore the clusters are *already* created (hence the calculations are a sixth step), to which the Examiners suggested amending the independent claims to more strongly clarify and positively recite that the cluster attractor is determined prior to the creation of a cluster. The Examiners were appreciative of the claim amendments in this regard of the last Response, but Primary Examiner PANNALA felt strongly that the claim terms “*for use in clustering*” and “*distribution that is indicative*” should be revised (with the former considered intended use, and not granted patentable weight).

Applicants’ Representatives and the Examiner also discussed independent method and apparatus claims 15 and 19, reciting “clustering a plurality of documents... comprising... select... a cluster attractor... compare each document with each cluster attractor; and... assign each document to one or more cluster attractors....” Although the Examiners favorably regarded the latter part of these claims, (Primary Examiner PANNALA said the latter recitation “should be alright”), they felt that this latter recitation did not resolve the perceived ambiguity of the recitation shared with independent claims 1 and 14, and thus these claims should be amended in similar fashion.

Primary Examiner PANNALA also strongly suggested that some kind of formula be recited in the claims, indicating that reciting unique formulas from the specification into the claims could help avoid an extensive amount of art. Examiner RUIZ indicated that the amendments of the Response filed September 30, 2009 overcame the objections to claims 1 and 18, and overcame the rejections under 35 U.S.C. §101. The Examiners stated that with the current claim language, Choi (Fig. 2, paragraphs [0027]-[0029]) reads on the broadest interpretation of the claims. Applicants’ Representatives and the Examiners further discussed possible amendments that would distinguish the

claimed subject matter from the applied references.

Applicants thank Examiner RUIZ for the Examiner Interview Summary mailed November 17, 2009 and the specific recommendations provided in the interview and the Summary to advance prosecution and distinguish the claimed subject matter. Applicants also thank the Examiner for her indication on December 14, 2009, that the amendments discussed above address the recommendations discussed with the Examiner(s) during the Interview. Accordingly, Applicants submit this Supplemental Amendment for the Examiner's review and consideration, and request that Applicants' Representatives be contacted should any additional issues remain outstanding preventing the application from allowance.

Claim Rejections under 35 U.S.C. §103(a)

The Examiner has rejected claims 1-4 and 15-18 under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent Application Publication No. 2002/0042793 to Choi (hereinafter referred to as "Choi") in view of U.S. Patent No. 4,839,853 to Deerwester et al. (hereinafter referred to as "Deerwester et al.") and rejected claims 5-12 and 14 as being unpatentable over Choi, in view of Deerwester et al. and further in view of U.S. Patent No. 6,128,613 to Wong et al. (hereinafter referred to as "Wong et al.").

Response

Claim 18 has been canceled without prejudice or disclaimer to the subject matter therein. Applicants respectfully traverse the remaining rejections since all of the features of the presently

claimed subject matter are not disclosed by the cited references. To establish a *prima facie* case of obviousness, the Examiner must establish: (1) some suggestion or motivation to modify the references exists; (2) a reasonable expectation of success; and (3) the prior art references teach or suggest all of the claim limitations. *Amgen, Inc. v. Chugai Pharm. Co.*, 18 USPQ2d 1016, 1023 (Fed. Cir. 1991); *In re Fine*, 5 USPQ2d 1596, 1598 (Fed. Cir. 1988); *In re Wilson*, 165 USPQ 494, 496 (CCPA 1970).

A *prima facie* case of obviousness must also include a showing of the reasons why it would be obvious to modify the references to produce the present invention. See *Dystar Textilsfarben GMBH v. C. H. Patrick*, 464 F.3d 1356 (Fed. Cir. 2006). The Examiner bears the initial burden to provide some convincing line of reasoning as to why the artisan would have found the claimed invention to have been obvious in light of the teachings. *Id.* at 1366.

Overview

Independent claim 1 recites “[a] computer-implemented method of determining cluster attractors for a plurality of documents, each document comprising at least one term, each term comprising one or more words, the method comprising:

causing a computer to calculate, in respect of each term, a probability distribution indicative of

in the instance where a document comprises said term and one other term that co-occurs with said term in at least one of said documents, the frequency of occurrence of said one other term, and

in the instance where a document comprises said term and more than one other term that co-occurs with said term in at least one of said documents, the respective frequency of occurrence of each other term;

causing a computer to calculate, in respect of each term, the entropy of the respective probability distribution; and

causing the computer to select at least one of said probability distributions as a cluster attractor depending on the respective entropy value,

wherein the selected cluster attractor is a clustering focus for at least some of said documents.”

These features are not described nor rendered obvious by Choi or Deerwester et al., whether taken alone or in combination. Claims 2-12 depend directly or indirectly from claim 1.

Independent claim 14 recites “[a]n apparatus for determining cluster attractors for a plurality of documents, each document comprising at least one term, each term comprising one or more words, the apparatus comprising:

means for calculating, in respect of each term, a probability distribution indicative of

in the instance where a document comprises said term and one other term that co-occurs with said term in at least one of said documents, the frequency of occurrence of said one other term, and

in the instance where a document comprises said term and more than one other term that co-occurs with said term in at least one of said documents, the respective frequency of occurrence of each other term;

means for calculating, in respect of each term, the entropy of the respective probability distribution; and

means for selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value,

wherein the selected cluster attractor is a clustering focus for at least some of said documents.”

These features are not described nor rendered obvious by Choi or Deerwester et al. or Wong et al., whether taken alone or in combination.

Independent claim 15 recites “[a] computer-implemented method of clustering a plurality of documents, each document comprising at least one term, each term comprising one or more words, the method comprising:

causing a computer to calculate, in respect of each term, a probability distribution

indicative of

in the instance where a document comprises said term and one other term that co-occurs with said term in at least one of said documents, the frequency of occurrence of said one other term, and

in the instance where a document comprises said term and more than one other term that co-occurs with said term in at least one of said documents, the respective frequency of occurrence of each other term;

causing a computer to calculate, in respect of each term, the entropy of the respective probability distribution;

causing the computer to select at least one of said probability distributions as a cluster attractor depending on the respective entropy value;

causing the computer to compare each document with each cluster attractor; and

causing the computer to assign each document to one or more cluster attractors depending on the similarity between the document and the cluster attractors,

wherein assigning each document to one or more cluster attractors creates a plurality of document clusters, each cluster comprising a respective plurality of documents.”

These features are not described nor rendered obvious by Choi or Deerwester et al., whether taken alone or in combination. Claims 16 and 17 depend directly from claim 15.

Choi relates to “[a] method of order-ranking document clusters using entropy data and Bayesian self-organizing feature maps(SOM) is provided in which an accuracy of information retrieval is improved by adopting Bayesian SOM for performing a real-time document clustering for relevant documents in accordance with a degree of semantic similarity between entropy data extracted using entropy value and user profiles and query words given by a user, wherein the Bayesian SOM is a combination of Bayesian statistical technique and Kohonen network that is a type of an unsupervised learning.” *See Choi Abstract.*

Deerwester et al. relates to “[a] methodology for retrieving textual data objects.... The information is treated in the statistical domain by presuming that there is an underlying, latent

semantic structure in the usage of words in the data objects. Estimates to this latent structure are utilized to represent and retrieve objects. A user query is recouched in the new statistical domain and then processed in the computer system to extract the underlying meaning to respond to the query.”

See Deerwester et al. Abstract.

Wong et al. relates to “[a] computer-based method and system for establishing topic words to represent a document, the topic words being suitable for use in document retrieval. The method includes determining document keywords from the document; classifying each of the document keywords into one of a plurality of preestablished keyword classes; and selecting words as the topic words, each selected word from a different one of the preestablished keyword classes, to minimize a cost function on proposed topic words. The cost function may be a metric of dissimilarity, such as cross-entropy, between a first distribution of likelihood of appearance by the plurality of document keywords in a typical document and a second distribution of likelihood of appearance by the plurality of document keywords in a typical document, the second distribution being approximated using proposed topic words. The cost function can be a basis for sorting the priority of the documents.” *See Wong et al. Abstract.*

Rejection of claims 1-4 and 15-18

Applicants respectfully submit that Deerwester et al. fails to cure the deficiencies of Choi with respect to the claimed subject matter in accordance with Applicants’ independent claims 1 and 15, and further, does not suggest a teaching or motivation to reach such subject matter as claimed in the instant application. Applicants further respectfully submit that the cited prior art of record does

not suggest a teaching or motivation to reach such subject matter as claimed in the instant application.

With regard to claim 1, Applicants disagree with the Examiner's assertion that Choi discloses "a... method of determining cluster attractors for a plurality of documents," as recited in the present claims. In contrast to the claimed subject matter, Choi does not use cluster attractors when clustering documents. Instead, Choi clusters documents by comparing documents to each other and putting documents that have a high similarity in the same cluster. *See* Choi paragraphs [0109] and [0110]; for example, "to group individual documents, a measure for clustering documents is needed. As a measurement, similarity and dissimilarity between documents is used. Here, if similarity between documents is employed as a measurement, documents having relatively higher similarity are classified into the same group. If dissimilarity is employed, documents having relatively lower dissimilarity are classified into the same group. The most fundamental method employing dissimilarity between documents is to use distance between documents" (Choi explains the specific manner in which he clusters documents by measuring their similarity/dissimilarity to one another in Choi paragraphs [0111]-[0121]).

The Examiner has cited Choi paragraph [0126], but this paragraph relates to computing the distance between clusters after the clusters have been formed. Therefore this paragraph does not relate to cluster attractors, which are used to form the clusters and accordingly must be calculated before the clusters are formed. The Examiner also refers to Choi paragraphs [0132] – [0134]. However, Choi merely describes the centroid linkage method which is a method of measuring the distance between two clusters that have already been formed – *see* for example Choi paragraph

[0133] “as a distance between the two clusters c1 and c2, the distance between the centroids of the two clusters is used” – i.e. the “centroid linkage method” referred to by Choi is performed after the clusters have been formed and therefore does not relate to cluster attractors, which are required before the clusters are formed. In this connection, the “centroid” mentioned by Choi in paragraphs [0133] and [0134] can only be calculated after the cluster is formed (see for example the equation given in paragraph [0133] in which centroids are calculated by using data relating to the documents in the clusters that have already been formed). Therefore, the centroid is not the same as a cluster attractor within the meaning of claim 1, since a cluster attractor is used to form the clusters and accordingly must be identified before the cluster is formed. To help clarify this distinction, Applicants have amended claim 1 to recite “...wherein the selected cluster attractor is a clustering focus for at least some of said documents.”

Applicants also respectfully disagree that Choi teaches to “calculate, in respect of each term, a probability distribution.” In this connection, the Examiner cited Choi paragraphs [0028] and [0029]. However, Choi paragraph [0029] describes “prior information” in “the form of probability distribution” and it can be seen from Choi paragraph [0027] that this “prior information” is an initial value for a neural network. An initial value for a neural network has nothing to do with the terms, e.g. words, of the document and so in contrast to that recited in claim 1, Choi does not describe that a probability distribution is calculated “in respect of each term.”

In particular, Choi does not describe the particular probability distribution recited in claim 1, namely a “*probability distribution indicative of[:]*” in the instance where a document comprises said term and one other term that co-occurs with said term in at least one of said documents, *the*

frequency of occurrence of said one other term, and in the instance where a document comprises said term and more than one other term that co-occurs with said term in at least one of said documents, *the respective frequency of occurrence of each other term*" (emphases added). As mentioned above, in contrast to claim 1, the probability distribution referred to by Choi in paragraph [0029] has nothing to do with the terms of the document and therefore has nothing to do with the frequency of occurrence and/or co-occurrence of terms within documents.

The Examiner has also cited Choi paragraph [0060] in connection with this part of claim 1. However, Choi paragraph [0060] simply states that documents can be clustered by grouping documents with similar contents into the same cluster by a document clustering technique. As discussed above, this may be regarded as *teaching away from* claim 1 since claim 1 is concerned with determining cluster attractors for clustering documents, whereas Choi, as demonstrated in paragraph [0060], clusters documents by comparing their similarity with each other.

Applicants also respectfully disagree that Choi teaches to "calculate, in respect of each term, the entropy of each of the respective probability distribution... select at least one of said probability distributions as a cluster attractor depending on the respective entropy value," recited in claim 1. In this connection, the Examiner cited Choi paragraphs [0051] to [0053]. It can be seen from Choi paragraph [0052] (and paragraph [0028]) that Choi only describes calculating an entropy value for "query words given by user profiles with respect of key words for each of the documents." Therefore, Choi does not teach to calculate the entropy value of a probability distribution of any type, and particularly not a probability distribution of the type recited in claim 1. Accordingly, Choi cannot "select at least one of said probability distributions as a cluster attractor depending on the respective

entropy.” Firstly, because he does not use cluster attractors (instead documents are compared directly with one another) and secondly, because he does not calculate entropy values of any probability distributions.

Therefore, it is respectfully submitted that claim 1 is patentable over Choi at least because of the above-discussed features.

With regard to Deerwester et al., Applicants respectfully submit that although Table 2 of Deerwester et al. describes nine technical document titles in a “term-by-document” matrix in which the frequency of occurrence of term in document j is recorded, Deerwester does not describe any of the features cited in claim 1. Firstly, Deerwester et al. is not concerned with document clustering and so does not require cluster attractors or describe any method of determining cluster attractors. Secondly, Deerwester et al. does not teach to “calculate, in respect of each term, a probability distribution indicative of in the instance where a document comprises said term and one other term that co-occurs with said term in at least one of said documents, the frequency of occurrence of said one other term, and in the instance where a document comprises said term and more than one other term that co-occurs with said term in at least one of said documents, the respective frequency of occurrence of each other term” as presently claimed. In this connection, it is noted in particular that Table 2 of Deerwester et al. describes actual frequencies of occurrence of the terms and documents and not probability distributions *indicative of the frequency of occurrence of said one other or each other term*, as per claim 1. Further, because Deerwester et al. does not teach probability distributions, it cannot teach to calculate the

entropy of the probability distributions. Also, because it does not use cluster attractors, Deerwester et al. cannot teach to “select at least one of said probability distributions as a cluster attractor depending on the respective entropy value, *wherein the selected cluster attractor is a clustering focus for at least some of said documents*” (emphases added) as presently claimed.

Accordingly, since neither Choi nor Deerwester et al. describe any of the above-discussed features of claim 1, it is respectfully submitted that their combined teachings would also not lead a person having ordinary skill in the art to subject matter of claim 1. Moreover, Choi and Deerwester et al. are not technically compatible with one another – Choi relates to document clustering while Deerwester et al. relates to retrieving textual data objects by user query by mapping all of the terms and documents from the document corpus into a semantic space. See Deerwester et al. col. 2, lines 24-40. Applicants respectfully submit that a person having ordinary skill in the art would not combine the teachings of these references, at least because they address entirely different issues. Therefore, it is respectfully submitted that claim 1 is not obvious over the combined teachings of Choi and Deerwester et al.

With regard to claim 15, claim 15 includes all of the features of claim 1 and has been amended to recite these explicitly. Therefore, it is respectfully submitted that claim 15 is novel and non-obvious over the individual combined teachings of Choi and Deerwester et al. for the reasons set out above in relation to claim 1.

Claim 15 additionally recites the features “*causing the computer to compare each document with each cluster attractor; and causing the computer to assign each document to one or more cluster attractors depending on the similarity between the document and the cluster attractors,*

wherein assigning each document to one or more cluster attractors creates a plurality of document clusters, each cluster comprising a respective plurality of documents” (emphases added). Choi does not describe these additional features of claim 15. This is because, as described above in relation to claim 1, Choi does not use cluster attractors when forming its clusters. Instead, and in contrast to the claimed subject matter, Choi forms clusters by comparing documents to each other (see paragraphs [0109] and [0110]), and any “order-ranking [of] document clusters using entropy data and Bayesian SOM” is performed only *after* formation of the clusters (*see* paragraphs [0027]-[0029], *sixth* and *seventh* steps).

In connection with claim 15, the Examiner has highlighted the Abstract of Choi. It is respectfully submitted however that the Choi Abstract has no relevance to claim 15 because claim 15 recites a computer implemented method of clustering a plurality of documents, i.e. claim 15 relates to the *formation* of clusters “wherein assigning each document to one or more cluster attractors *creates a plurality of document clusters, each cluster comprising a respective plurality of documents*” (emphases added). In contrast, the Choi Abstract describes “a method of order ranking document clusters”, i.e. the Choi Abstract relates only to what happens after the clusters are formed. As indicated previously, Choi describes a method of clustering, but this is achieved by comparing the similarities/dissimilarities between documents rather than determining cluster attractors in the manner defined in claim 15 and then comparing the documents to the cluster attractors as per claim 15. It is respectfully submitted that claim 15 is novel and non-obvious over the individual and combined teachings of Choi and Deerwester et al.

The dependent claims are patentable at least due to their dependency from patentable independent claims 1 and 15. Applicants refer to the Response filed September 30, 2009 for further discussion regarding the features of the dependent claims not addressed above.

Rejection of claims 5-12 and 14

Applicants respectfully submit that Wong et al. fails to cure the deficiencies of Choi and Deerwester et al. with respect to the claimed subject matter in accordance with Applicants' independent claims 1 and 14, and further, does not suggest a teaching or motivation to reach such subject matter as claimed in the instant application. Applicants further respectfully submit that the cited prior art of record does not suggest a teaching or motivation to reach such subject matter as claimed in the instant application. The dependent claims are patentable at least due to their dependency from patentable independent claim 1.

With regard to independent claim 14, claim 14 is an apparatus claim reciting, *inter alia*, features corresponding to the features discussed above with respect to method claim 1. Applicants therefore submit that claim 14 is novel, non-obvious and consequently patentable over Choi and Deerwester et al. (and Wong et al.) for the reasons discussed above in relation to claim 1.

Applicants respectfully submit that the cited art, whether taken alone or in combination, fails to describe, teach or suggest each and every feature of the claimed subject matter as recited in the independent claims. The dependent claims are patentable at least due to their dependency from patentable independent claims. Applicants refer to the Response filed September 30, 2009 for further

discussion regarding the patentable features of the dependent claims. Applicants respectfully request withdrawal of the rejections of claims 1-12 and 14-17.

Amended claim 19

Amended claim 19 is an apparatus claim reciting, *inter alia*, features corresponding to the features discussed above with respect to method claim 15. Applicants respectfully submit that claim 19 is novel, non-obvious and consequently patentable over Choi and Deerwester et al. (and Wong et al.) for the reasons discussed above in relation to claim 15.

Newly presented claims 20-21

Newly presented claims 20 and 21 are method claims reciting, *inter alia*, features corresponding to the features discussed above with respect to method claim 1. Claims 20 and 21 additionally recite, as suggested by the Examiner(s), specific formulae for calculating the probability distribution and entropy, respectively. The subject matter of newly presented independent claim 20 is supported throughout the specification, claims and figures as originally filed, at least by original and amended claim 1 and page 14, line 1, et seq. of the original specification. The subject matter of newly presented dependent claim 21 is similarly supported, and at least by original and amended claim 1 and page 15, line 11, et seq. of the original specification. It is respectfully submitted that the amendments do not introduce any new matter within the meaning of 35 U.S.C. §132. Applicants respectfully submit that claims 20 and 21 are novel, non-obvious and consequently patentable over Choi and Deerwester et al. for the reasons discussed above in relation to claim 1.

CONCLUSION

In light of the foregoing, Applicants submit that the application is in condition for allowance.

If the Examiner believes the application is not in condition for allowance, Applicants respectfully request that the Examiner call the undersigned.

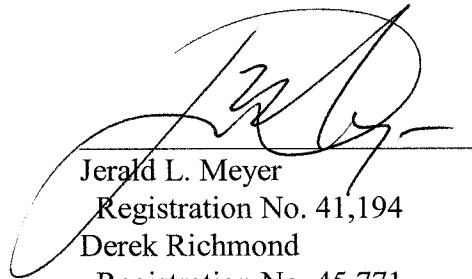
In the event this paper is not timely filed, Applicants hereby petition for an appropriate extension of time. Please charge any fee deficiency or credit any overpayment to Deposit Account No. 14-0112.

Respectfully submitted,

THE NATH LAW GROUP

December 14, 2009

THE NATH LAW GROUP
112 South West Street
Alexandria, VA 22314-2891
Tel: 703-548-6284
Fax: 703-683-8396



Jerald L. Meyer
Registration No. 41,194

Derek Richmond
Registration No. 45,771

Jiaxiao Zhang
Registration No. 63,235
Customer No. 20529